

PATENT ABSTRACTS OF JAPAN

(11) Publication number: **01191258 A**(43) Date of publication of application: **01.08.89**

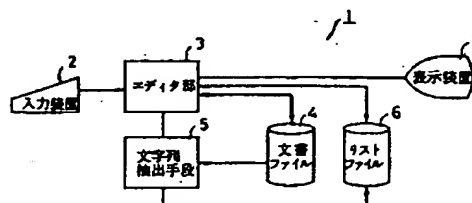
(51) Int. Cl.

G06F 15/20(21) Application number: **63015316**(71) Applicant: **RICOH CO LTD**(22) Date of filing: **26.01.88**(72) Inventor: **NARITA MASUMI****(54) DOCUMENT PROCESSING DEVICE****(57) Abstract:**

PURPOSE: To quickly and accurately perform the extracting work by extracting character strings generated in accordance with a prescribed sequence, for example, reference literature names from a prescribed document and storing them.

CONSTITUTION: A prescribed document, for example, the text of a treatise is stored in a document file 4 from an input device 2, and reference literature names cited in this text are automatically extracted by a character string extracting means 5 and are listed up in a list file 6. Thereafter, the text of the treatise stored in the document file 4 and reference literature names listed up in the list file 6 are edited in an editor part 3 to complete one treatise. Thus, prescribed character strings of reference literature names or the like are quickly and accurately extracted from the document.

COPYRIGHT: (C)1989,JPO&Japio



⑨ 日本国特許庁(JP)

⑩ 特許出願公開

⑫ 公開特許公報(A) 平1-191258

⑮ Int. Cl.⁴

識別記号

庁内整理番号

⑬ 公開 平成1年(1989)8月1日

G 06 F 15/20

3 0 1

V-7218-5B

審査請求 未請求 請求項の数 1 (全5頁)

⑭ 発明の名称 文書処理装置

⑯ 特 願 昭63-15316

⑰ 出 願 昭63(1988)1月26日

⑱ 発 明 者 成 田 真 澄 東京都大田区中馬込1丁目3番6号 株式会社リコー内

⑲ 出 願 人 株 式 会 社 リ コ ー 東京都大田区中馬込1丁目3番6号

明 細 書

1. 発明の名称

文書処理装置

2. 特許請求の範囲

入力された文字列を文書として記憶する入力記憶手段と、所定のシーケンスで生起する文字列を前記入力記憶手段に記憶された所定の文書から抽出する文字列抽出手段と、抽出された文字列を格納する格納手段とを備えていることを特徴とする文書処理装置。

3. 発明の詳細な説明

〔技術分野〕

本発明は、機械翻訳装置、ワードプロセッサなどに利用される文書処理装置に関し、より詳細には、文書中から所定の文字列、例えば参考文献名を抽出する文書処理装置に関するものである。

〔従来の技術〕

一般に、文書例えば英文で論文を作成する際には、論文中に引用された参考文献名を論文の最後にリストアップするようになっている。

従来、このような作業は、本文を書き終えたあと参考文献名を本文の冒頭から手作業で抜き出すというように非能率的なものとなっており、また手作業のため見落とし等があるという問題があった。

〔目的〕

本発明は、文書中から参考文献名などの所定の文字列を迅速にかつ正確に抽出することの可能な文書処理装置を提供することを目的としている。

〔構成〕

本発明は上記の目的を達成させるため、入力された文字列を文書として記憶する入力記憶手段と、所定のシーケンスで生起する文字列を前記入力記憶手段に記憶された所定の文書から抽出する文字列抽出手段と、抽出された文字列を格納する格納手段とを備えていることを特徴としたものである。

以下、本発明の一実施例に基づいて説明する。

第1図は本発明の文書処理装置の構成図である。この文書処理装置1は、キーボードなどの入力装置2と、文字列を編集するエディタ部3と、エディタ部3で編集された入力装置2からの文字列を文書として格納する文書ファイル4と、文書ファイル4に格納された所定の文書から所定の文字列、例えば参考文献名を抽出する文字列抽出手段5と、抽出された所定の文字列を格納する格納手段すなわちリストファイル6と、文書ファイル4に格納された所定の文書、およびリストファイル6に格納された所定の文字列をエディタ部3を介し表示するディスプレイなどの表示装置7とを備えている。

このような文書処理装置1では、文書ファイル4に所定の文書、例えば論文の本文を入力装置2から格納し、文字列抽出手段5により例えば論文の本文中に引用されている参考文献名を自動的に抽出し、これをリストファイル6にリストアップし、しかる後、文書ファイル4に格納された論文

の本文とリストファイル6にリストアップされた参考文献名とをエディタ部3で編集し、1つの論文として完成させるようになっている。

第2図は論文の本文中に引用されている参考文献名の文字列構造を説明するための図であり、第2図の例では参考文献名の文字列は、大文字で始まる英略文字列 "Hayes" と、開き括弧 "(" と、数字文字列 "1968" と、閉じ括弧 ")" とが順次にシーケンスとして生起するものとして定義されている。

第3図乃至第6図は、論文中の参考文献名が第2図に示すような文字列構造をもつとしたときの文書処理装置1、特に文字列抽出手段5の処理流れを示すフローチャートであり、第3図はメインルーチン、第4図、第5図、第6図はそれぞれ第3図のメインルーチンに用いられる文字保存領域の初期化処理、文字保存処理、リストファイルへの書き込み処理のサブルーチンである。

第3図においてステップS1、S2では、文書ファイル4を開き、続いてリストファイル6を開

く。

次いでステップS3では、文字列の識別用変数STに初期値 "0" を設定する。なお識別用変数STは、文字列のシーケンスを表示する領域であって、第2図に示すように大文字 "H"、アルファベット "ayes"、開き括弧 "(", 数字 "1968" が順次に生起すると "1", "2", "3", "4" のように変化するものである。このようにして識別用変数STの初期化を終了するとステップS4において文字列抽出手段5自体に備わっている文字列保存領域(図示せず)の初期化を行なう。

文字列保存領域の初期化は、第4図のようにして行なわれる。すなわち、ステップS40でポインタIに "1" を設定し、ステップS41で文字列保存領域CH(I)をクリアする。次いでステップS42でポインタIが既定値 "N"、例えば "20" となったか否か判断し、"N" でないときにはステップS43でポインタIを "1" だけ歩進させ再びステップS41に戻し、ポインタI

が "N" となるまでCH(I)をクリアする。ポインタIが "N" になったときには、文字列保存領域CH(1)~CH(N)がクリアされ、初期化処理を終了し、第3図のステップS5に進む。

ステップS5では、ポインタIを "0" に初期化する。次いでステップS6では文書ファイル4から1文字読み、ステップS7では文書の終わりで文字が読出されたか否かを判断する。文書の終わりで文字が読出されたときにはステップS8、S9に進み文書ファイル4並びにリストファイル6を閉じて処理を終了する。ステップS7において文書の終わりで読出されたものでなければ、ステップS10に進み、読出された文字が大文字か否かを調べる。大文字であればステップS11に進み識別用変数STを "1" に設定し、次いでステップS12において文字保存処理を行なう。

この文字保存処理は第5図に示すようにして行なわれ、ステップS50でポインタIを "1" だけ歩進し、ステップS51において文字列保存領域CH(I)に読出された文字を格納する。いま

の場合、ステップS5においてポインタ1は“0”となっているので、文字列保存領域CH(1)に読出された大文字、例えば“H”が格納されることになる。

ステップS12において文字保存処理を行なった後、再びステップS6に戻り、次の文字を文書ファイル4から読出す。

ステップS10において読出された文字が大文字でないときには、ステップS13に進み、この文字がアルファベットか否かを判断する。アルファベットであるときにはステップS14に進み、識別用変数STが“1”または“2”であるかを判断する。識別用変数STが“1”となっているときには、大文字に続いてアルファベットが生起したことを意味し、また識別用変数STが“2”となっているときには、大文字、アルファベットに続いてアルファベットの文字が再び読出されたことを意味するので、第2図の参考文献名の文字列の候補となり、ステップS15に進み識別用変数STを“2”に設定し、ステップS16の文字

文字を読出す。

ステップS17において開き括弧“(”でないかと判断されたときには、ステップS21に進み、読出された文字が数字か否かを調べる。数字のときには、ステップS22に進み、識別用変数STが“3”または“4”であるかを調べる。“3”または“4”のときには、大文字、アルファベット、開き括弧に続いて数字が生起したかまたは大文字、アルファベット、開き括弧、数字に続いて数字が生起したことを意味するので、参考文献名の候補となり、ステップS23に進んで識別用変数STに“4”を設定し、ステップS24の文字保存処理を行なって再びステップS6に戻り次の文字を読出す。

一方、ステップS22において識別用変数STが“3”でも“4”でもないときには、参考文献名の候補とはなり得ないので、ステップS3に戻る。

またステップS21において読出された文字が数字でないときには、ステップS25に進み、こ

保存処理に進み、次の文字列保存領域CH(1)に読出された文字を格納し、再びステップS6に戻って次の文字を読出す。

一方、ステップS14において識別用変数STが“1”でも“2”でもないときには、大文字、アルファベット以外の文字に続いてアルファベットの文字が読出されたことを意味するので、参考文献名の文字列の候補とはなり得ず、ステップS3の処理に戻る。

また、ステップS13において読出された文字がアルファベットでなければステップS17に進み、開き括弧“(”か否かが調べられる。開き括弧“(”であれば、ステップS18に進み識別用変数STが“2”であるか否かを判断する。“2”のときには、大文字、アルファベットに続いて開き括弧“(”が読出されたことを意味するので、ステップS19に進み識別用変数STに“3”を設定し、ステップS20において文字保存処理を行ない、次の文字列保存領域CH(1)に読出された文字を格納して再びステップS6に戻り次の

の文字が閉じ括弧”)”であるか否かを調べる。閉じ括弧”)”であるときには、ステップS26に進み、識別用変数STが“4”であるか否かを調べる。“4”のときには、大文字、アルファベット、開き括弧、数字に続いて閉じ括弧”)”が生起し参考文献名の文字列であることを意味するので、ステップS27に進み文字保存処理を行なう。これにより1つの参考文献名が抽出され、文字列保存領域CH(1)には参考文献名を構成する文字列が順次に格納されたので、ステップS28においてこの文字列保存領域CH(1)の内容をリストファイル6へ順次に書き込み、再びステップS3に戻る。

なお、リストファイル6への書き込み処理は、第6図のようにして行なわれる。すなわち、ステップS60ではポインタ1の値をポインタJに移し、ステップS61でポインタ1に“1”を設定し、ステップS62ではポインタ1でアドレスされる文字列保存領域CH(1)の内容をリストファイル6に書き込む。次いでステップS63では、ポイ

ンタ1とポインタJとを比較し、同じでないときにはステップS64に進みポインタ1を“1”だけ歩進させ、再びステップS62に戻る。ステップS63においてポインタ1とポインタJとが同じになったときには、文字列保存領域CH(1)に保存されている参考文献名の文字列が全てリストファイル6へ書き込まれたことを意味するので書き込処理を終了する。

またステップS25において閉じ括弧“)”でないかと判断されたときには、ステップS29に進み、識別用変数STが“4”であるか否かを調べる。“4”でないときには、参考文献名の候補とはならないのでステップS3に戻るが、“4”であるときには、入力エラーにより閉じ括弧“)”が抜けるという可能性があるため、ステップS30に進み入力エラーにより閉じ括弧が抜けたか否かをチェックするチェック処理を行ない、入力エラーと判明したときには文字列保存領域CH(1)に順次に格納された文字列は参考文献名を構成すると判断され、ステップS28に進み、文

字列保存領域CH(1)の内容をリストファイル6へ順次に書き込み、ステップS3に戻る。

ステップS3に戻ったとき、文書ファイル4の所定の文書から再び次の参考文献名を抽出する処理が行なわれ、この繰返し処理は、文書の終わりで文字が読出されるまで続く。

このようにして、上記繰返し処理が終了したときに、リストファイル6には、抽出した所定の文字列すなわち参考文献名が全て格納される。

なお、上述の処理では、参考文献名ではないが参考文献名と同じ構造をしている文字列、例えば“Figure(3)”も抽出される可能性がある。しかし、このような過剰抽出はリストファイル6をエディタ部3によりあとで編集する際に削除すれば良く、抽出もれが生ずるよりもはるかに問題が少ない。

〔効果〕

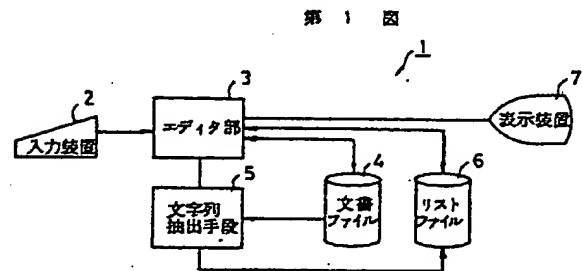
以上に説明したように、本発明によれば、所定のシーケンスで生起する文字列、例えば参考文献名を所定の文書から抽出し、格納するようになっ

ているので、手作業で抽出する場合に比べて、迅速にかつ正確に行なうことができる。

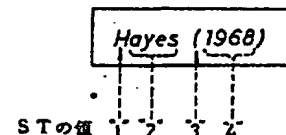
4. 図面の簡単な説明

第1図は本発明の文書処理装置の構成図、第2図は抽出されるべき文字列の構造を説明するための図、第3図は文書処理装置の処理流れを示すフローチャート、第4図乃至第6図はそれぞれ文書保存領域の初期化処理、文字保存処理、リストファイルへの書き込処理のフローチャートである。

- 1…文書処理装置、2…入力装置、
- 3…エディタ部、4…文書ファイル、
- 5…文字列抽出手段、6…リストファイル、
- 7…表示装置



第2図



特許出願人 株式会社 リ コ ー

